(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/GB02/04809

(22) International Filing Date: 24 October 2002 (24.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0126436.5   3 November 2001 (03.11.2001)   GB

(71) Applicant (for all designated States except US): SE-QUENOM-GEMINI LIMITED [GB/GB]; Cartyle House, 107 Wellington Road South, Stockport SK1 3TL (GB).

(72) Inventors; and
(75) Inventors/Applicants (for US only): WILSON, Scott, Geoffrey [AU/AU]; 6 Southern Cross Circle, Ocean Reef, W.A. 6027 (AU). SALLSTROM, Eva, Gunvor, Kristina [SE/SE]; Svartbacksgatan 103, Uppsala, DS 753 35 (SE). BANSAL, Aruna [GB/GB]; 13 High Street, Landbeach, Cambridge CB4 8DR (GB). REED, Peter [NZ/GB]; 21 Glebe Way, Haddenham, Cambridge CB6 3TG (GB).

(74) Agent: THOMSON, P., A.; Potts, Kerr & Co., 15 Hamilton Square, Birkenhead CH41 6BR (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DETECTION OF PREDISPOSITION TO OSTEOPOROSIS

(57) Abstract: The invention provides novel reagents, kits, and methods for diagnosis of predisposition to low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis, based on analysis of polymorphic variants of the nucleic acid set forth in SEQ ID NO:1.

DETECTION OF PREDISPOSITION TO OSTEOPOROSIS

The present invention relates to oligonucleotides, kits, microarrays, and methods for detection of bone disease, in particular, osteopenia and osteoporosis.

BACKGROUND OF THE INVENTION

5   Osteoporosis is a disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and consequently increased bone fracture. The cost of osteoporotic fracture in the US alone is estimated at $13.8 billion per annum. Fracture is the clinical endpoint in osteoporosis, but bone mineral density (BMD) is commonly used as a surrogate for determining risk of fracture. BMD is

10   an estimate of the mineral mass, corrected for the area (anteroposterior projection) of the bone under study. BMD is the strongest predictor of osteoporotic fracture known, and this measurement is made using Dual X-Ray Absorptiometry (DEXA). Data derived from DEXA scans of the lumbar spine (L1-L4) and hip (total hip and femoral neck) can be used in studies of osteoporosis.

15   In women, peak bone mass is reached between age 20-30, after which there is a period of consolidation and gradual loss. However, there is a rapid decline in BMD of approximately 3% per year for approximately 5 years after menopause. Consequently, estrogen is strongly implicated as a crucial element in the maintenance of bone mass. This is supported by the role of hormone replacement therapy in the prevention and treatment of

20   osteoporosis. A role for estrogen has also been implicated in the growth and maintenance of bone in men.

Twin and family resemblance studies show that osteoporosis has a substantial genetic component. Data shows that bone mineral density (BMD) has a heritability in the range of 0.66 to 0.82. Other data on Colles' (wrist) fracture in families, also supports the genetic

25   basis of osteoporotic fracture. This data shows that the relative risk of Colles' fracture, if a woman's mother or sisters have had a Colles' fracture, is 1.3 and 1.9 respectively. On the basis of this, and other published data, osteoporosis is regarded as a complex multifactorial disease with multiple gene and environmental factors contributing to disease susceptibility.

Lifestyle changes such as moderate exercise; ensuring sufficient calcium intake, limiting alcohol intake; and cessation of smoking, are indicated for individuals with osteopenia and osteoporotic bone disease. Osteopenia and osteoporosis can be treated using lifestyle changes, hormone replacement therapy (HRT: estrogen), selective estrogen receptor

5   modulators (SERMs), antiresorptive agents (bisphosphonates), calcitonin and sodium fluoride. The optimal treatment for osteoporosis has not been determined.

A number of genes have been identified which are associated with predisposition to osteoporosis. For example, U.S. Pat. No. 5,922,542 discloses an association between risk of developing osteoporosis and certain polymorphisms in the gene encoding collagen 1α1,

10   also referred to as COL1A1, and a kit based on this association has been commercialized in Europe.

U.S. Pat. No. 5,998,137 discloses methods of diagnosing a number of diseases, including osteoporosis, by detecting a polymorphism in the promoter of the transforming growth factor β (TGF-β) gene. WO 00/23618 discloses a method of detecting predisposition to

15   osteoporosis on the basis of a polymorphism in intron 5 of the TGF-β gene. JP2000270897 also discloses a method for detection of danger of osteoporosis based on a polymorphism in the TGF-β gene.

U.S. Pat. No. 5,698,399 discloses methods of diagnosing predisposition to osteoporosis by detecting certain polymorphic variants in the gene encoding interleukin-1 receptor

20   antagonist.

U.S. Pat. No. 6,066,450 discloses methods for detecting predisposition to osteoporosis by identifying certain polymorphic variants in the interleukin-6 gene.

EP 1054066 discloses a method for determining sensitivity to an osteoporosis medication based on analysis of certain polymorphisms in the genes encoding vitamin D receptor, the

25   estrogen receptor, and apolipoprotein E.

WO 01/09383 suggests that polymorphisms in the human gene encoding the melatonin-related receptor, a G-protein coupled receptor of unknown function, may be involved in bone-related disorders, including osteoporosis.

WO 01/23559 suggests that mutations in the regulatory region of the human gene encoding

30   osteoclast differentiation factor may be used to detect or predict susceptibility to bone diseases, including osteoporosis.

WO 01/20031 discloses that certain single nucleotide polymorphisms (SNPs) in the human *klotho* gene are associated with forearm and spine bone mineral density and thus may be used to diagnose predisposition to osteoporosis.

Because osteoporosis is likely to be the result of genetic variations in multiple genes,
5 additional genes and polymorphisms which may be associated with this condition are the subject of ongoing research. A need remains for identification of such genes, to enhance physicians' ability to detect osteopenia and predisposition to osteoporosis in individuals who may have the condition, but do not exhibit symptoms. Thus lifestyle changes or therapy can be initiated early in the progression of disease.

10   **SUMMARY OF THE INVENTION**

On the basis of genetic analysis of dizygotic twins and sib pairs, the present inventors have discovered that certain polymorphisms in the nucleic acid set forth in SEQ ID NO:1 are associated with variation in bone mineral density (BMD). Specifically, the presence of an A nucleotide at position 245 of SEQ ID NO:1 is associated with low spine BMD, low total
15 hip BMD, and low femoral neck BMD, and the presence of a C nucleotide at position 245 of SEQ ID NO:1 is associated with higher spine, total hip, and femoral neck BMD. In addition, the presence of a G nucleotide at position 1470 of SEQ ID NO:1 is associated with low total hip and femoral neck BMD, and the presence of an A nucleotide at position 1470 of SEQ ID NO:1 is associated with higher total hip and femoral neck BMD.
20 Moreover, the presence of a haplotype represented by an A nucleotide at position 245 of SEQ ID NO:1 and a G nucleotide at position 1470 of SEQ ID NO:1 is associated with low spine BMD, low total hip BMD, and low femoral neck BMD. These associations may be used as the basis of reagents, kits, and methods for detection of predisposition to osteoporosis.
25 The nucleic acid of SEQ ID NO:1 is a draft human genomic sequence of nucleotides 58007412 to 58016140 of chromosome 3 (Human Genome Project Working Draft, University of California, Santa Cruz, April 2001 Freeze), which comprises a cDNA (GenBank Accession No. XM_003213) corresponding to a gene of unknown function currently known as E2IG3. The complete genomic sequence of E2IG3 is available in a
30 chromosome 3 working draft sequence with GenBank Accession No. NT_005986 [gi:

3

15297785]. .E2IG3 is upregulated by 17β estradiol (estrogen). The protein encoded by E2IG3 is believed to belong to a subfamily of large GTP-binding proteins. The E2IG3 cDNA is also disclosed as SEQ ID NO:247 of WO 00/55350 and as SEQ ID NO: 6137 of WO 00/58473.

5   In one embodiment, the invention provides a sequence determination oligonucleotide complementary to a polymorphic region within a nucleic acid having a sequence as set forth in SEQ ID NO:1, wherein the region corresponds to a polymorphic site selected from the group consisting of position 245 of SEQ ID NO:1 and position 1470 of SEQ ID NO:1.

10   In another-embodiment, the invention provides a microarray comprising at least one oligonucleotide complementary to a polymorphic region in the nucleic acid set forth in SEQ ID NO:1, wherein the region corresponds to a polymorphic site selected from the group consisting of position 245 of SEQ ID NO:1 and position 1470 of SEQ ID NO:1.

In another embodiment, the invention provides the oligonucleotide primer pairs useful for

15   amplification of a polymorphic region in the nucleic acid of SEQ ID NO:1 from a biological sample, wherein the region corresponds to a polymorphic site selected from the group consisting of position 245 of SEQ ID NO:1 and position 1470 of SEQ ID NO:1.

In another embodiment, the invention provides a kit comprising at least one

20   oligonucleotide primer pair complementary to a polymorphic region of the nucleic acid of SEQ ID NO:1, wherein the region corresponds to a polymorphic site selected from the group consisting of position 245 of SEQ ID NO:1 and position 1470 of SEQ ID NO:1.

The invention is also embodied in a method of diagnosing predisposition to low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral

25   density, or osteoporosis in a human, said method comprising the steps of obtaining a nucleic acid sample from the human; detecting the presence or absence of at least one allelic variant of a polymorphic region in a nucleic acid having a sequence as set forth in SEQ ID NO:1 in the sample, wherein the polymorphic region corresponds to the polymorphic site at position 245 of SEQ ID NO:1.

30   The invention is also embodied in a method of diagnosing predisposition to low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis in a

4

human, said method comprising the steps of obtaining a nucleic acid sample from the

human; and detecting the presence or absence of at least one allelic variant of a

polymorphic region in a nucleic acid having a sequence as set forth in SEQ ID NO:1 in

the sample, wherein the polymorphic region corresponds to the polymorphic site at

5   position 1470 of SEQ ID NO:1.

In a further embodiment, the invention provides a method of diagnosing predisposition to

low spine bone mineral density, low total hip bone mineral density, low femoral neck

neck bone mineral density, or osteoporosis in a human comprising the steps of obtaining

a nucleic acid sample from the human; and detecting the presence or absence of a

10   haplotype of the nucleic acid having a sequence as set forth in SEQ ID NO:1, said

haplotype being characterized by: an A nucleotide at position 245 of SEQ ID NO:1 and a

G nucleotide at position 1470 of SEQ ID NO:1.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 sets forth the nucleic acid of SEQ ID NO:1 with the polymorphic sites at

15   positions 245 and 1470 in bold capital type. The transcription start site of E2IG3 is

indicated in italics, introns are depicted in lower case type, and exons are depicted in bold

lower case type. Exons are as described in GenBank Accession No. XM_003213.

Figure 2 sets forth the sequences of certain oligonucleotides of the invention, which are

correlated with the polymorphic site the oligonucleotides are designed to detect.

20   Polymorphic sites in these oligonucleotides are indicated in bold capital type.

Figure 3 sets forth the sequences of certain oligonucleotide primer pairs designed to

amplify polymorphic regions of the nucleic acid of SEQ ID NO:1.

DETAILED DESCRIPTION OF THE INVENTION

The U.S. patents and publications referenced herein are hereby incorporated by reference.

25   Examples 1 and 2 below demonstrate associations between certain polymorphic regions

in SEQ ID NO:1 and measurements of bone mineral density.

For the purposes of the invention, certain terms are defined as follows.

"Oligonucleotide" means a nucleic acid molecule preferably comprising from about 8 to

about 50 covalently linked nucleotides. More preferably, an oligonucleotide of the

invention comprises from about 8 to about 35 nucleotides. Most preferably, an
oligonucleotide of the invention comprises from about 10 to about 25 nucleotides. In
accordance with the invention, the nucleotides within an oligonucleotide may be analogs
or derivatives of naturally occurring nucleotides, so long as oligonucleotides containing

5   such analogs or derivatives retain the ability to hybridize specifically within the
polymorphic region containing the targeted polymorphism. Analogs and derivatives of
naturally occurring oligonucleotides within the scope of the present invention are
exemplified in U.S. Pat. Nos. 4,469,863; 5,536,821; 5,541,306; 5,637,683; 5,637,684;
5,700,922; 5,717,083; 5,719,262; 5,739,308; 5,773,601; 5,886,165; 5,929,226; 5,977,296;

10   6,140,482; WO 00/56746; WO 01/14398, and the like. Methods for synthesizing
oligonucleotides comprising such analogs or derivatives are disclosed, for example, in the
patent publications cited above and in U.S. Pat. Nos. 5,614,622; 5,739,314; 5,955,599;
5,962,674; 6,117,992; in WO 00/75372, and the like. The term "oligonucleotides" as
defined herein includes compounds which comprise the specific oligonucleotides

15   disclosed herein, covalently linked to a second moiety. The second moiety may be an
additional nucleotide sequence, for example, a tail sequence such as a polyadenosine tail
or an adaptor sequence, for example, the phage M13 universal tail sequence, and the like.
Alternatively, the second moiety may be a non-nucleotidic moiety, for example, a moiety
which facilitates linkage to a solid support or a label to facilitate detection of the

20   oligonucleotide. Such labels include, without limitation, a radioactive label, a fluorescent
label, a chemiluminescent label, a paramagnetic label, and the like. The second moiety
may be attached to any position of the specific oligonucleotide, so long as the
oligonucleotide retains its ability to hybridize to the polymorphic regions described
herein.

25   A polymorphic region as defined herein is a portion of a genetic locus that is
characterized by at least one polymorphic site. A genetic locus is a location on a
chromosome which is associated with a gene, a physical feature, or a phenotypic trait. A
polymorphic site is a position within a genetic locus at which at least two alternative
sequences have been observed in a population. A polymorphic region as defined herein

30   is said to "correspond to" a polymorphic site, that is, the region may be adjacent to the
polymorphic site on the 5' side of the site or on the 3' side of the site, or alternatively may

contain the polymorphic site. A polymorphic region includes both the sense and antisense strands of the nucleic acid comprising the polymorphic site, and may have a length of from about 100 to about 5000 base pairs. For example, a polymorphic region may be all or a portion of a regulatory region such as a promoter, 5′ UTR, 3′ UTR, an

5   intron, an exon, or the like. A polymorphic or allelic variant is a genomic DNA, cDNA, mRNA or polypeptide having a nucleotide or amino acid sequence that comprises a polymorphism. A polymorphism is a sequence variation observed at a polymorphic site, including nucleotide substitutions (single nucleotide polymorphisms or SNPs), insertions, deletions, and microsatellites. Polymorphisms may or may not result in detectable

10 differences in gene expression, protein structure, or protein function. Preferably, a polymorphic region of the present invention has a length of about 1000 base pairs. More preferably, a polymorphic region of the invention has a length of about 500 base pairs. Most preferably, a polymorphic region of the invention has a length of about 200 base pairs.

15 A haplotype as defined herein is a representation of the combination of polymorphic variants in a defined region within a genetic locus on one of the chromosomes in a chromosome pair. A genotype as used herein is a representation of the polymorphic variants present at a polymorphic site.

A polymorphic region of the present invention comprises a portion of SEQ ID NO:1

20 corresponding to at least one of the polymorphic sites identified above. That is, a polymorphic region of the invention may include a nucleotide sequence surrounding and/or including any of the polymorphic sites at positions 245 and 1470 of SEQ ID NO:1. Polymorphic regions in the antisense nucleic acid complementary to SEQ ID NO:1 are also encompassed in the present invention, wherein the region includes a nucleotide

25 sequence surrounding and/or including any of the antisense positions corresponding to positions 245 and 1470 of SEQ ID NO:1. Figure 2 sets forth exemplary oligonucleotides within the scope of this embodiment. For example, a polymorphic region corresponding to the polymorphic site at position 245 of SEQ ID NO:1 may comprise a sequence as set forth in any of SEQ ID NO:2; SEQ ID NO:3; SEQ ID NO:4; SEQ ID NO:5; SEQ ID

30 NO:6; SEQ ID NO:7; SEQ ID NO:8; or SEQ ID NO:9. A polymorphic region corresponding to the polymorphic site at position 1470 of SEQ ID NO:1 may comprise a

7

sequence as set forth in any of SEQ ID NO:10; SEQ ID NO:11; SEQ ID NO:12; SEQ ID
NO:13; SEQ ID NO:14; SEQ ID NO:15; SEQ ID NO:16; or SEQ ID NO:17.

In certain embodiments of the invention, oligonucleotides are used as probes for the
polymorphic regions in the nucleic acid having the sequence set forth in SEQ ID NO:1.

5   These oligonucleotides may also be termed "sequence determination oligonucleotides"
within the scope of the invention, and may be used to determine the presence or absence
of a particular nucleotide at a particular polymorphic site within the nucleic acid of SEQ
ID NO:1. Specific oligonucleotides of the invention include any oligonucleotide
complementary to any of the polymorphic regions described above.

10   Those of ordinary skill will recognize that oligonucleotides complementary to the
polymorphic regions described herein must be capable of hybridizing to the polymorphic
regions under conditions of stringency such as those employed in primer extension-based
sequence determination methods, restriction site analysis, nucleic acid amplification
methods, ligase-based sequencing methods, methods based on enzymatic detection of

15   mismatches, microarray-based sequence determination methods, and the like. The
oligonucleotides of the invention may be synthesized using known methods and
machines, such as the ABI™3900 High Throughput DNA Synthesizer and the
EXPEDITE™ 8909 Nucleic Acid Synthesizer, both of which are available from Applied
Biosystems (Foster City,CA).

20   The oligonucleotides of the invention may be used, without limitation, as *in situ*
hybridization probes or as components of diagnostic assays. Numerous oligonucleotide-
based diagnostic assays are known. For example, primer extension-based nucleic acid
sequence detection methods are disclosed in U.S.Pat.Nos. 4,656,127; 4,851,331;
5,679,524; 5,834,189; 5,876,934; 5,908,755; 5,912,118; 5,976,802; 5,981,186; 6,004,744;

25   6,013,431; 6,017,702; 6,046,005; 6,087,095; 6,210,891; WO 01/20039; and the like.
Primer extension-based nucleic acid sequence detection methods using mass
spectrometry are described in U.S. Pat. Nos. 5,547,835; 5,605,798; 5,691,141; 5,849,542;
5,869,242; 5,928,906; 6,043,031; 6,194,144, and the like. The oligonucleotides of the
invention are also suitable for use in ligase-based sequence determination methods such

30   as those disclosed in U.S.Pat.Nos. 5,679,524 and 5,952,174, WO 01/27326, and the like.
The oligonucleotides of the invention may be used as probes in sequence determination

8

methods based on mismatches, such as the methods described in U.S.Pat.Nos. 5,851,770; 5,958,692; 6,110,684; 6,183,958; and the like. In addition, the oligonucleotides of the invention may be used in hybridization-based diagnostic assays such as those described in U.S.Pat.Nos. 5,891,625; 6,013,499; and the like.

5   The oligonucleotides of the invention may also be used as components of a diagnostic microarray. Methods of making and using oligonucleotide microarrays suitable for diagnostic use are disclosed in U.S. Pat. Nos. 5,492,806; 5,525,464; 5,589,330; 5,695,940; 5,849,483; 6,018,041; 6,045,996; 6,136,541; 6,142,681; 6,156,501; 6,197,506; 6,223,127; 6,225,625; 6,229,911; 6,239,273; WO 00/52625; WO 01/25485; WO

10   01/29259; and the like. Preferably, the microarray of the invention comprises at least one oligonucleotide complementary to a polymorphic region of SEQ ID NO:1, wherein the region corresponds to a polymorphic site selected from the group consisting of position 245 of SEQ ID NO:1 and position 1470 of SEQ ID NO:1. More preferably, the microarray of the invention comprises an oligonucleotide complementary to a

15   polymorphic region corresponding to position 245 of SEQ ID NO:1 and an oligonucleotide complementary to a polymorphic region corresponding to position 1470 of SEQ ID NO:1. In a specific embodiment, the oligonucleotides of the microarray of the invention are complementary to any or all of the polymorphic regions selected from the group consisting of SEQ ID NO:2; SEQ ID NO:3; SEQ ID NO:4; SEQ ID NO:5; SEQ ID

20   NO:6; SEQ ID NO:7; SEQ ID NO:8; and SEQ ID NO:9 (corresponding to the polymorphic site at position 245 of SEQ ID NO:1); and the polymorphic regions selected from the group consisting of SEQ ID NO:10; SEQ ID NO:11; SEQ ID NO:12; SEQ ID NO:13; SEQ ID NO:14; SEQ ID NO:15; SEQ ID NO:16; and SEQ ID NO:17 (corresponding to the polymorphic site at position 1470 of SEQ ID NO:1).

25   The invention is also embodied in oligonucleotide primer pairs suitable for use in the polymerase chain reaction (PCR) or in other nucleic acid amplification methods. Each oligonucleotide primer pair of the invention is complementary to a polymorphic region of the nucleic acid of SEQ ID NO:1. Thus an oligonucleotide primer pair of the invention is complementary to a polymorphic region characteristic of at least one of the

30   polymorphic sites at positions 245 and 1470 of SEQ ID NO:1. Those of ordinary skill will be able to design suitable oligonucleotide primer pairs using knowledge readily

available in the art, in combination with the teachings herein. Specific oligonucleotide
primer pairs of this embodiment include the oligonucleotide primer pairs set forth in SEQ
ID NO:18 and SEQ ID NO:19, which are suitable for amplifying the polymorphic region
corresponding to the polymorphic site at position 245 of SEQ ID NO:1; and the

5   oligonucleotide primer pairs set forth in SEQ ID NO:20 and SEQ ID NO:21, which are
suitable for amplifying the polymorphic region corresponding to the polymorphic site at
position 1470 of SEQ ID NO:1. Those of skill will recognize that other oligonucleotide
primer pairs suitable for amplifying the polymorphic regions of the nucleic acid of SEQ
ID NO:1 can be designed without undue experimentation. In particular, oligonucleotide

10   primer pairs suitable for amplification of larger portions of SEQ ID NO:1 would be
preferred for haplotype analysis.
Each of the PCR primer pairs of the invention may be used in any PCR method. For
example, a PCR primer pair of the invention may be used in the methods disclosed in
U.S.Pat.Nos. 4,683,195; 4,683,202, 4,965,188; 5,656,493; 5,998,143; 6,140,054; WO

15   01/27327; WO 01/27329; and the like. The PCR pairs of the invention may also be used
in any of the commercially available machines that perform PCR, such as any of the
GENEAMP® Systems available from Applied Biosystems.
The invention is also embodied in a kit comprising at least one oligonucleotide primer
pair of the invention. Preferably, the kit of the invention comprises at least two

20   oligonucleotide primer pair, wherein each primer pair is complementary to a different
polymorphic region of the nucleic acid of SEQ ID NO:1. More preferably, the kit of the
invention comprises at least one oligonucleotide primer pair suitable for amplification of
polymorphic regions corresponding to positions 245 or 1470 of SEQ ID NO:1. This
embodiment may optionally further comprise a sequence determination oligonucleotide

25   for detecting a polymorphic variant at any or all of the polymorphic sites corresponding
to positions 245 and 1470 of SEQ ID NO:1. The kit of the invention may also comprise
a polymerizing agent, for example, a thermostable nucleic acid polymerase such as those
disclosed in U.S. Pat. Nos. 4,889,818; 6,077,664, and the like. The kit of the invention
may also comprise chain elongating nucleotides, such as dATP, dTTP, dGTP, dCTP, and

30   dITP, including analogs of dATP, dTTP, dGTP, dCTP and dITP, so long as such analogs
are substrates for a thermostable nucleic acid polymerase and can be incorporated into a

10

growing nucleic acid chain. The kit of the invention may also include chain terminating nucleotides such as ddATP, ddTTP, ddGTP, ddCTP, and the like. In a preferred embodiment, the kit of the invention comprises at least one oligonucleotide primer pair, a polymerizing agent, chain elongating nucleotides, at least one sequence determination

5   oligonucleotide and at least one chain terminating nucleotide. The kit of the invention may optionally include buffers, vials, microtiter plates, and instructions for use.

Methods of diagnosing predisposition to low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis in a human are also encompassed by the present invention. In the methods of the invention, the

10   presence or absence of at least one polymorphic variant of the nucleic acid of SEQ ID NO:1 is detected to determine or diagnose such a predisposition. Specifically, in a first step, a nucleic acid is isolated from biological sample obtained from the human. Any nucleic-acid containing biological sample from the human is an appropriate source of nucleic acid for use in the methods of the invention. For example, nucleic acid can be

15   isolated from blood, saliva, sputum, urine, cell scrapings, biopsy tissue, and the like. In a second step, the nucleic acid is assayed for the presence or absence of at least one allelic variant of any or all of the polymorphic regions of the nucleic acid of SEQ ID NO:1 described above. Preferably, the polymorphic regions on both chromosomes in the chromosome pair of the human are assayed in the method of the invention, so that the

20   zygosity of the individual for the particular polymorphic variant may be determined. Any method may be used to assay the nucleic acid, that is, to determine the sequence of the polymorphic region, in this step of the invention. For example, any of the primer extension-based methods, ligase-based sequence determination methods, mismatch-based sequence determination methods, or microarray-based sequence determination methods

25   described above may be used, in accordance with the present invention. Alternatively, such methods as restriction fragment length polymorphism (RFLP) detection, single strand conformation polymorphism detection (SSCP), PCR-based assays such as the TAQMAN® PCR System (Applied Biosystems) may be used.

In accordance with one method of the invention, predisposition to low spine bone mineral

30   density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis is diagnosed by determining the identity of the nucleotide at position 245 of

SEQ ID NO:1. In this method, an A nucleotide at position 245 of SEQ ID NO:1, or a T nucleotide at the corresponding position of the antisense complement of SEQ ID NO:1, is indicative of greater risk of developing low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis.

5   Conversely, the presence of a C nucleotide at position 245 of SEQ ID NO:1, or of a G nucleotide at the corresponding position of the antisense complement of SEQ ID NO:1, is indicative of a lower risk of developing low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis. In a further step, the zygosity of the individual may be determined, wherein a homozygous

10   AA genotype at position 245 of SEQ ID NO:1 or TT genotype at the corresponding position of the antisense complement of SEQ ID NO:1, indicates greatest risk for developing low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis. A person whose genotype is homozygous CC at position 245 of SEQ ID NO:1 or GG at the corresponding position of

15   the antisense complement of SEQ ID NO:1 is at least risk for developing low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis. An individual whose genotype is heterozygous AC at position 245 of SEQ ID NO:1 or TG at the corresponding position of the antisense complement of SEQ ID NO:1 is at intermediate risk for developing low spine bone mineral density, low

20   total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis. Alternatively, predisposition to low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis is diagnosed by determining the identity of the nucleotide at position 1470 of SEQ ID NO:1. In this embodiment, predisposition to low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis is

25   diagnosed by determining the identity of the nucleotide at position 1470 of SEQ ID NO:1. In this method, a G nucleotide at position 1470 of SEQ ID NO:1, or a C nucleotide at the corresponding position of the antisense complement of SEQ ID NO:1, is indicative of greater risk of developing low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis. Conversely, the presence of an A nucleotide

30   at position 1470 of SEQ ID NO:1, or of a T nucleotide at the corresponding position of the antisense complement of SEQ ID NO:1, is indicative of a lower risk of developing

12

low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis. In a further step, the zygosity of the individual may be determined, wherein a homozygous GG genotype at position 1470 of SEQ ID NO:1 or CC genotype at the corresponding position of the antisense complement of SEQ ID NO:1, indicates greatest

5   risk for developing low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis. A person whose genotype is homozygous AA at position 1470 of SEQ ID NO:1 or TT at the corresponding position of the antisense complement of SEQ ID NO:1 is at least risk for developing low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis. An individual whose genotype is

10   heterozygous GA at position 1470 of SEQ ID NO:1 or CT at the corresponding position of the antisense complement of SEQ ID NO:1 is at intermediate risk for developing low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis. In another method of the invention, risk of low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis is assessed

15   by determining the haplotype of the individual for both polymorphic positions within SEQ ID NO:1. For example, individuals who possess the SEQ ID NO:1 haplotype characterized by an A nucleotide at position 245 of SEQ ID NO:1 and a G nucleotide at position 1470 of SEQ ID NO:1 are at higher risk of development of low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral

20   density, or osteoporosis. This haplotype may alternatively be detected on the antisense complement of SEQ ID NO:1 as a T nucleotide at the antisense position corresponding to position 245 of SEQ ID NO:1 and a C nucleotide at the antisense position corresponding to position 1470 of SEQ ID NO:1. Individuals who are homozygous for allelic variants comprising this haplotype are at particularly high risk of developing low spine bone

25   mineral density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis.

The examples set forth below are provided as illustration and are not intended to limit the scope and spirit of the invention as specifically embodied therein.

EXAMPLE 1

Clinical samples

A study design based on extreme discordant and concordant sib pairs (EDACS) was chosen for analysis. All available female sibs of families containing EDACS were
5   included in the sample. EDACS were defined on the criteria of BMD Z score. In each family, probands required a BMD Z for total spine (L1-4), total hip or femoral neck of Z $\leq$ -1.5. At least one additional sib in the family was required to have either BMD for total spine (L1-4), total hip or femoral neck equal to Z $\leq$ -1.0 or Z $\geq$ 1.0. Z scores were derived from individual BMD scan data (Hologic) of the subjects and transformed using
10   the published Hologic reference range for spine BMD (Favus, 1999 in *Primer on the metabolic bone diseases and disorders of mineral metabolism*, Ed. Favus, MJ, 4th Edition, Lippincott Williams and Wilkins, Philadelphia, USA. 1999:483-484 or NHANESIII (Looker *et al.* (1995) *J. Bone Mineral Res.* **10**, 796-802) for hip and femoral neck. Any BMD data collected using Lunar technology was transformed to the Hologic BMD
15   equivalent using sBMD (Genant *et al.* (1994) *J. Bone Mineral Res.* **9**:1503-14; Steiger *et al.* (1995) *J. Bone Mineral Res.* **10**:1602; Hanson *et al.* (1997) *J. Bone Mineral Res.* **12**:1316).

A BMD Z score < −1.5 or > +1.5 corresponds to the bottom or top 6.7% of the age matched distribution, respectively. A Z score of −1.0 or +1.0 corresponds to being in the bottom or top 15.9% of the age matched distribution and includes both osteoporotic and osteopenic individuals. A total of 1098 samples were selected and plated for analysis as

5   the study sample set.

Whole genome scans were performed on 1401 twin pairs using up to 706 highly polymorphic microsatellite markers (Reed et al. (1994) Nat Genet 7:390-5). Additional whole genome scans were performed on 649 subjects from 283 families containing a proband with low bone mineral density. These later subjects were genome scanned using

10   the Affymetrix HuSNP GENECHIP® technology platform (Wilson et al. (2000) Calcified Tissue International 67:484). Multipoint nonparametric linkage analysis was performed using MAPMAKER/SIBS (Kruglyak & Lander (1995) Amer. J. Human Genetics 57:439-454).

For fine mapping and SNP genotyping a study design based on extreme exclusion criteria

15   for probands and sibs were used and verified by questionnaire if possible. These exclusion criteria were myeloma, osteosarcoma, or malignancy with skeletal involvement, hyperparathyroidism, unstable thyroid disease, long term steroid use (>5 mg/day for more than 6 months and presently on therapy), chronic immobility, rheumatoid arthritis, anorexia nervosa (>1 yr), history of osteomalacia, amenorrhea for >

20   6 months, premature cessation of regular menstruation or surgical oophorectomy +/- /HRT (age <35 yrs), very late menarche >20years of age, 15>BMI>40, and epilepsy with use of anticonvulsant medication for >1 year. These exclusions were applied conservatively and only subjects who had substantial evidence that they should be excluded were removed, because it was not possible to verify the exact details with the

25   patients or the medical records.

EXAMPLE 2

Genotyping

The linkage studies used a proprietary bioinformatics infrastructure and proprietary software packages to record marker positions, store data and generate data files (See WO

30   00/51053). Output from these systems was then used with the relevant application software to perform the statistical analysis.

MAPMAKER/SIBS (Kruglyak & Lander (1995) *Amer. J. Human Genetics* 57, 439-454) was used to estimate multipoint nonparametric linkage for the fine mapping studies. The computer program QTDT (Abecasis *et al.* (2000) *Amer. J. Human Genetics* 66, 279-292) was used to test for evidence of association. Three tests were considered:

5   Transmission/disequilibrium test (TDT), population stratification and total association tests. Population stratification can lead to spurious results from total association testing, so the latter result was considered and reported only when there was no evidence of population stratification.

Haplotype analysis was with QPDT (Martin *et al.* (2000) *Amer. J. Human Genetics* 67,

10   146-54), which utilises the EM (Dempster *et al.* (1977) *J. Royal Statistical Soc.* B39, 1-38) algorithm to assign haplotypes based on likelihood maximization.

A.   **Microsatellite Fine Mapping**

Thirty-five microsatellite markers in a broad interval (50.4 cM – 111 cM) on

15   chromosome 3 were chosen for analysis in the study population, with a mean spacing of 2.02 cM between each marker, across the region. Genotyping reactions were generally carried out in microtitre plates (384-well, reaction volume 5µl), containing 12.5ng of DNA DNA from study subjects was amplified using PCR and sequence specific oligonucleotide primers labelled with 6-FAM™, HEX™, or NED™ fluorescent dyes.

20   PCR products were analysed by electrophoresis in a polyacrylamide denaturing gel, with an ABI PRISM™ GENESCAN® 400HD ROX labelled size standard in each lane on an ABI model 377 analyzer (Applied Biosystems, Foster City, California). For genotyping, the chosen markers were divided into two groups (panels) so that the analysis of all of the markers could be performed in two electrophoresis runs of each sample. Consequently,

25   there was no overlap of fragment sizes in any one dye for either of the panels. Genotype analysis was performed using ABI PRISM™ GENESCAN® software (version 3.0), and genotyped manually using ABI PRISM™ Genotyper 2.0. Results were input into a proprietary database and binned by marker. The results were quality checked, ensuring consistent inheritance within families. Families that were found to have consistent

30   pedigree problems were excluded from the analysis set.

The ordering of genetic mapping markers (i.e. microsatellite markers) was relatively stable in the region analyzed according to the Unified Data Base for Human Genome Mapping, Weizmann Institute of Science (UDB) and National Center for Biotechnology Information, National Institutes of Health (NCBI) assemblies during the duration of the study. Conversion of genetic to physical positions for strategic microsatellite markers was performed using UDB and NCBI as the reference standards. Comparisons of the identity and positioning of genomic contigs in the region were also made between UDB and NCBI and provided relatively good agreement. A comparison of the positioning of all identified and predicted genes within the region was also made between NCBI (build 22) and Joint Project between European Bioinformatics Institute and the Sanger Centre (ENSEMBL). At its broadest the region encompassed genomic contigs NT_005980.3 to NT_005607.3 (build 22) or NT_005498.4 to NT_005589.4 (build 24). Given the identified mapping inconsistencies between public domain genome assemblies, several additional contigs were also considered. The major focus within this broad region was between NT_006022.3 and NT_ 005589.3 (NCBI build 22; 43.9Mb – 57.2Mb).

The microsatellite marker analysis showed linkage of spine BMD Z to chromosome 3p21 with a non parametric $Z = 4.07$ at 68.8 cM (n=1619 pedigrees). Using the -1 LOD approach, the support interval was 62.16 cM-75.62 cM. Total hip gave a peak Z score within the region of 1.17 at 65.1 cM and femoral neck $Z = 1.02$ at 60 cM.

The additional markers increased the information content for linkage, from a mean of 0.4382 to 0.5623, (range of 0.3177-0.6967). In the region of greatest interest, from 50 cM to 100 cM, information content for linkage has mean 0.6007 and range 0.5357-0.6967.

No association was found using multi-allelic test of association for microsatellite markers within the support interval.

B. SNP Genotyping.

SNPs analysed in the study were sourced from the public SNP resource, dbSNP, which contains an estimated 1.4 million non-redundant SNPs mapped to the NCBI genome browser (Stoneking, 2001, *Nature* 409: 821-822).

SNPs were selected at approximately 50 kb intervals across the region analyzed. The following attributes were taken into account when selecting SNPs: amount of available 5'

17

and 3' sequence context, absence of repeat masking in surrounding sequence, presence of multiple submissions, identity and reputation of the submitter. Relatively high SNP failure rate has been predicted, so several rounds of SNP selection were planned and undertaken.

5   Where SNP validation and screening were performed, the SNP assay was run on a sample set composed of 96 samples and replicated in quadruplicate. Results were examined to verify a working PCR reaction and appropriate PCR product, and to establish that some level of polymorphism existed.

Five SNPs in the E2IG3 gene (also known as Q9UJY0) reported by dbSNP were

10   examined. Two of the SNPs (corresponding to position 245 of SEQ ID NO:1, and position 1470 of SEQ ID NO:1) were genotyped in the complete sample set. The SNPs were amplified (GENEAMP™ PCR system 9700, Applied Biosystems) using 4.6 ng of genomic DNA in a total volume of 22.3 µl per well. Genotyping was performed using PSQ™ 96 SNP Reagents Kit 5x96 and SNP detection was subsequently performed using

15   the PSQ 96 platform (Pyrosequencing AB, Uppsala, Sweden). In order to detect any genotyping anomalies, Hardy-Weinberg, haplotype analysis and duplicate control genotypings were performed. No deviation in genotyping results could be seen in the duplicates. The frequency of the three different genotypes for each SNP did not differ significantly from expected values in the Hardy-Weinberg test. Both markers and were

20   found to show significant association by TDT with total hip BMD Z ($p=0.0009$ and $p=0.0045$ respectively). Data for femoral neck BMD Z also shows significant association with the two markers ($p=0.0424$ and $p=0.0272$ respectively). In comparison, spine BMD showed only a weak association to spine BMD with p-values 0.058 and 0.061 respectively, although haplotypes analysis did show association ($p= 0.0325$ for haplotype

25   AG). Analysis with QPDT gave similar results with the lowest p-value for association being with the SNP corresponding to position 245 of SEQ ID NO:1 and total hip BMD, a p-value of 0.0002. The association with spine BMD was 0.016 and 0.076, respectively. Weight is known to explain a substantial proportion of the variance in BMD at weight bearing sites, and particularly at the hip. Consequently, weight adjusted hip BMD was

30   also studied in relation to the E2IG3 association with hip BMD. After adjustment for

18

weight, evidence of association with total hip BMD remains in the analysis (e.g. position 245 of SEQ ID NO:1; p= 0.003).

As a result of these findings fragments covering exon 2, 12, 13 and 14 of E2IG3 were sequenced using solid phase sequencing (AUTOLOAD™ Solid Phase Sequencing kit,

5   Amersham Pharmacia Biotech) and gel electrophoresis on ALFEXPRESS™ sequencers (Amersham Pharmacia Biotech) to identify additional SNPs, between the associated SNPs and in the 3' end of the gene. One additional SNP has been discovered in Exon 13. This is a G to C polymorphism at position 7840 in SEQ ID NO:1. Preliminary data suggests that the frequency of this SNP is below 5%.

10

While the invention has been described in terms of the specific embodiments set forth above, those of skill will recognize that the essential features of the invention may be varied without undue experimentation and that such variations are within the scope of the appended claims.

15

CLAIMS

1.   A microarray comprising at least one oligonucleotide complementary to a polymorphic region within a nucleic acid having a sequence as set forth in SEQ ID NO:1, wherein the region corresponds to a polymorphic site selected from the group consisting of position 245 of SEQ ID NO:1 and position 1470 of SEQ ID NO:1.

2.   The microarray of claim 1, comprising an oligonucleotide complementary to a polymorphic region corresponding to position 245 of SEQ ID NO:1 and an oligonucleotide complementary to a polymorphic region corresponding to position 1470 of SEQ ID NO:1.

3.   An oligonucleotide complementary to a polymorphic region within a nucleic acid having a sequence as set forth in SEQ ID NO:1, wherein the region corresponds to a polymorphic site selected from the group consisting of position 245 of SEQ ID NO:1 and position 1470 of SEQ ID NO:1.

4.   The oligonucleotide of claim 3, wherein the region comprises a sequence selected from the group consisting of SEQ ID NO:2; SEQ ID NO:3; SEQ ID NO:4; SEQ ID NO:5; SEQ ID NO:6; SEQ ID NO:7; SEQ ID NO:8; SEQ ID NO:9; SEQ ID NO:10; SEQ ID NO:11; SEQ ID NO:12; SEQ ID NO:13; SEQ ID NO:14; SEQ ID NO:15; SEQ ID NO:16; and SEQ ID NO:17.

5.   A pair of oligonucleotide primers for amplifying a polymorphic region in a nucleic acid having a sequence as set forth in SEQ ID NO:1 from a biological sample, wherein the region corresponds to a polymorphic site selected from the group consisting of position 245 of SEQ ID NO:1 and position 1470 of SEQ ID NO:1

6.   The primers of claim 5, having sequences selected from the group consisting of: SEQ ID NO:18 and SEQ ID NO:19; and SEQ ID NO:20 and SEQ ID NO:21.

7.   A kit comprising at least one oligonucleotide primer pair complementary to a polymorphic region of a nucleic acid having a sequence as set forth in SEQ ID NO:1, wherein the region corresponds to a polymorphic site selected from the group consisting of position 245 of SEQ ID NO:1 and position 1470 of SEQ ID NO:1.

5

8.   The kit of claim 7, comprising at least two oligonucleotide primer pairs, wherein each primer pair is complementary to a different polymorphic region of the nucleic acid of SEQ ID NO:1.

10   9.   A method of diagnosing predisposition to low spine bone mineral density, low total hip bone mineral density, low femoral neck bone mineral density, or osteoporosis in a human, said method comprising the steps of:
obtaining a nucleic acid sample from the human; and
detecting the presence or absence of at least one allelic variant of a polymorphic region in
15   a nucleic acid having a sequence as set forth in SEQ ID NO:1 in the sample,
wherein the polymorphic region corresponds to the polymorphic site at position 245 of SEQ ID NO:1.

10.   A method of diagnosing predisposition to low total hip bone mineral density, low
20   femoral neck bone mineral density, or osteoporosis in a human, said method comprising the steps of:
obtaining a nucleic acid sample from the human; and
detecting the presence or absence of at least one allelic variant of a polymorphic region in
a nucleic acid having a sequence as set forth in SEQ ID NO:1 in the sample,
25   wherein the polymorphic region corresponds to the polymorphic site at position 1470 of SEQ ID NO:1.

30

21

1/10

FIGURE 1


```
ctgctgctat tgctcctccg gccgcggccg ctgccgtcgc ttcggcaccc gccgccctca
60

cctcccttac ccctcccggt gccgccgcaa aaccagtccc gcggccgcca agcgatccct
120

gctccgcgcg acactgcgtg cccgcgcacg cagagaggcg gtgacgcact ttacggcggc
180

agcgtaagtg cgtgacgctc gtcagtggct tcagttcaca cgtggcgcca gcggaggcag
240

gttgA/Ctgtgt ttgtgcttcc ttctacagcc aatatgaaaa ggcctagtaa gtggggtcgg
300

gaggcgggcg tggagggacc cacgtctgga agttgctgca gccaccacga cgctcttcta
360

cggctacggc tttgtctctg ctggtatggg ggtgggagcc tacgcgtagg ccttggccct
420

atttcctggt agaaccgaga gttggaagtc cctacggcga tcatgttaac cgcgcgggct
480

cattctgcgg aacgaagccg ggcagagggt ggggaagact aggctagatt ttcgtaagga
540

agcagcgtct gagccaggtt tgaggcccaa tattttcttt ccgtggccac gtgcagactg
600

gcccaggtga gagctgagaa tcgcctccca gactcagtgt tcctctcctg ccttatgatt
660

cgtgctgttt gacacgaagt ggttgtcgtt ttgtgtctca tacgctgttg tgtatgatcc
720

cattctaata ttgtgagggt aagtgcaggg aattttgact ccattctgga tctactgaat
780

ttaattctct gggatttgaa agtagcacgt atgtttgcat taggcatttc gcattagact
840

taacgttagg tttggtagcc aataacacaa gaaaaggata taactccata gtgcgttaac
900

ccagaactaa tcatttgggt taacagattt gtgatgtgtt tctttgtaga gttaaagaaa
960

gcaagtaaac gcatgacctg ccataagcgg tataaaatcc aaaaaaaggt aagtgtagtg
1020

cttgagagag ctgtaccaaa cacattgcta aactgatttt gccctgttcc tttgcgggaa .
1080
```

```
agtctgggtt aatgtgattt ggttttggga aatggcattg gatagactga ccatgggcac
1140

aagctcttag gcatcaggag tgcagctgtg agaaagtgca gtgatttggt gataagtctc
1200

taaatttgtt cagcatgtta atctctgcat agagagcctt ctagttacaa tttcttgctg
1260

ttttatactt acatatgcat tactttgtaa gattccaatt aaagctccat tttcctagga
1320

cattttatag gcataactaa attgcagcca gattggtttc tcacttgaat tctgcttaag
1380

tataaagata tttttgtaag cagacaaaat ctctttattt taataggttc gagaacatca
1440

tcgaaaatta agaaaggagg ctaaaaagcG/A gggtcacaag aagcctagga aagacccagg
1500

agttccaaac agtgctccct ttaaggaggc tcttcttagg gaagctgagc taaggaaaca
1560

gagggtaagt tatgttagcc agaattttca ttgagtggtg tagtgtgtta tgtgtgatat
1620

ttttcagagt aaggtaacaa cactagtcac tggttcacct atttccctta tggctctgac
1680

agcttgaaga actaaaacag cagcagaaac ttgacaggca gaaggaacta gaaaagaaaa
1740

gaaaacttga aactaatcct gatattaagc catcaaatgt ggaacctatg gaaaaggtat
1800

gattaggtct ctttatgaat gagagatcag gggttttgat tttggttttt ttgcttgggc
1860

ctgagtgcag tggcacaatc acaactcact gcaacctgga ccacccaggc tcaagcagtc
1920

ttaccacctc agtctccaag tagctgggac tacagaagca caccatcacg cctggctaat
1980

tttttttagc agacacgggc tttcactata ttgccaaggc tggtctcaag tgatccaccc
2040

aactcagcct cccgaagttt ctggtattac aggtgtgggc tgttgtgcct ggctgagaga
2100

tgagttctga tgcagaaata aaagcacatc cacaggctgc tgagcttctt gggaggaaga
2160

caactgagtt cagactccat cttacctatt taacaactgc aagggctgct actcagctgt
2220
```

ggaaaatgga gttagaggtt acagttgctc tacttctaat tttgtgttat ttcccccttt
2280

atcctctagg agtttgggct ttgcaaaact gagaacaaag ccaagtcggg caaacagaat
2340

tcaaagaagc tgtactgcca agaacttaaa aaggtatctt agcctaggtc agtgtctgac
2400

agtagtaatg aggtttaaaa gactcaagtc attttttttt taacctttta agatgaaggg
2460

tgcatgtgca ggtttgttat atgggtaaac ttgcgtcata ggggtttgct gtacagatta
2520

tttcatcacc caggtattaa gcctagtacg acattagtta tctttcctga tcctctccct
2580

cctcccacct tttcaccaac aaatcatttc gtgactcgac tctagcttat gctgtttaat
2640

gcctttcctg ctatgtttac ctgacggaaa tagtttcttt ggttctaaat atttgcacaa
2700

aactggttct gcctgtaagc atgatttaca acattaaaaa aaaacgttga catagtgttg
2760

agattgagaa aggtacattg gagtaagcag tgtcaggcta aaggtctcta aagtactctg
2820

ttgaaaccta agtgaaggag gacaacttgg tgtagttgct ctccagcact cccatcccca
2880

acatccattt tcccaagctc actcccccat ggatagaacc tgactgcccc tcagcagctt
2940

ttggcaaggc cagaaggacc tatcaaacta tataattcac tatgggagga ttcagacagg
3000

gatatttgca tttttgaaat ccatcttgat cagagactgc tgagcaagcc tatgttttac
3060

tttcctgtgt gagaaatgat gagggtcaac attcttcata ccaaagtgaa gacatgagat
3120

ccaactctga gctcaccctg ttgctaaatg gataatgcca gtactctctt gtggaaggta
3180

ttaccagaac aagggatgta gttctgatca ttttctcctt gataatgtag ttctggtcat
3240

tttttccttg ataggtgatt gaagcctccg atgttgtcct agaggtgttg gatgccagag
3300

atcctcttgg ttgcagatgt cctcaggtag aagaggccat tgtccagagt ggacagaaaa
3360

```
agctggtact tatattaaat aaatcaggtg agtaaagagg gtacccttttg tcttctgtgt
3420

acatgggtga ggtacgagga aacagtctga tagtcactga agactgatta gatccaactc
3480

tgatctcagc aaagccagag tacgtgcact ttgccagaga cagtgctagg cagtggggag
3540

ccaggtgact tttacaactg actcaactgg tttctactat tcttttgcca ttcagtattt
3600

accatctttt aaataaagag tgtaagctgc tatacccagc ttattgtgta gtatatttca
3660

tctaggaagt gatgacagtg tgacaaattc cccacaccta cacaatgtcg ggtattagtt
3720

caagagtgaa ataaattgga acgtatgtga caaaatattt aaatgaaatg cataattatg
3780

catctgagtt tgagcagcag gaaaaagaaa acccagaaca gagaattaca aagcagaaaa
3840

tgggaatgag atctaaaatt gttgttgggg ttaagaaaca attggctgct tgggaggctg
3900

aagtgggcac atcacttgag gccaggagtt cgagaaaagc ctggccaaca cggcgaaacc
3960

ccatctctac taaaatacaa atattagccg ggcatgatga tgggcacttg tagtcccagc
4020

tactcggaag gctgaggcag gagaattgct tgaacccggg aggcggaggt tgcagtgagc
4080

cgatactgtg ccactgcact ccagcctggg caacaacact tcgtctcaaa aatgaagaaa
4140

caattggctg ctagctaaag gtaaattctg gaaacatagc tctagggtta gtagggttgt
4200

aatccagacg ttgagtctgt cctgaagttt tcaagtgagc aatacaaggg gaattgaaat
4260

agagaagtgc agatgctgag ctctgttaag atacgggcag tatggtaggg gagcttaccc
4320

tgccctgatt ttctagttaa atccttttga aaggactggg aaaatgtaaa ccagagtaaa
4380

atctatagtt gccagatttt gcaaatgcat ctcaacaaaa tagccacatt ggagcaaatg
4440

tcttttttctt tttttctttt tgagatggag gcttgctgtg tcacccaggc tggagtgcag
4500
```

tggcgccatc tcagctcact gcaagctcca cctcccgggt tcacgccgtt ctcctgcctc
4560

agcctcccga gtagctggga ctacaggtgc ccaccaccac gcccagctaa tttttttgta
4620

tttttagtag agacggggtt ttatcgtgtt agccaggatg gtcttgatct cctgaactcg
4680

tgatccgcct gtctcggcct cccaaagtgc tgggattaca ggcgtgagcc accgtgcctg
4740

gccgcaaatg tcttatttct aattgctatc agatctggta ccaaaggaga atttggagag
4800

ctggctaaat tatttgaaga aagaattgcc aacagtggtg ttcagagcct caacaaaacc
4860

aaaggataaa gggaagataa ccaaggtatc ctttattagt ggtaagaaat gtgattcttt
4920

cagattttgg ttgaaatatg atgagtgtac aaaatcttga tttaagtgaa tgaaaaatta
4980

caagatccaa ctctgatttc agccagagat catctgaaag gcaatgtagt tatcttaaga
5040

gctgggctct ggagcctgat tgcttggggt ttgttgaaat ttatcaggta agttgccaga
5100

ataattcaac tatttggaat tttagcgtgt gaaggcaaag aagaatgctg ctccattcag
5160

aagtgaagtc tgctttggga aagagggcct ttggaaactt cttggaggtt ttcaggaaac
5220

ttgcagcaaa gccattcggg ttggagtaat tggtgagttt cagttcatta ctttttactt
5280

tttaagtgtt gaaatagtta agaagtttag ccagctctcc aagtgcccaa gcagcagtgt
5340

atggagttgt tgtcaagtaa aaggctcact caaatactag cttcttgctt ataccttact
5400

gaaagcacat aaccacacac aatttaaaga aaaaaactta cacaactgct gccagattca
5460

ggttttttgt tgggactgat tactgtagga agctggtttc taaaagttct tggtttgttt
5520

gaatttatag tattttcctg cctttgcatt acttgtgcaa gaaatgaaga aactaaaatt
5580

ggtcttagta ttgaagtgaa gacactgaga tccaactctg atcttgccct aaacatcagg
5640

```
gaaatggaaa attaggcagt gaaaatttca taagtgccaa catataatgc tttttatatc
5700

tggatttccc atttatttgt aggtttccca aatgtgggga aaagcagcat tatcaatagc
5760

ttaaaacaag aacagatgtg taatgttggt gtatccatgg ggcttacaag gtaaatggag
5820

gtgtccataa ttgtaatatt atagtgacac actattttat tttggttatc tcaaggaagg
5880

tgattttttt tttttttttt tttgagacag tttcactctt gttcccaggc tggagggcaa
5940

tggcgcaatc tcggctcact gcaacctcca cctctcaggt tcaagtgatt ctcctgcctc
6000

agcctccaga gtagctggga ttacaggcat gtgccccac gcccggctaa ttttgtattt
6060

ttaatagaga cgggtttctc catgttggtt aggctggtct caaactcctg acttcagctg
6120

atctgcccgc ctcggcctcc caaagtgctg gggttacagg ccaagccacc gcgcccggcc
6180

tatttttatt ttttaaagct tttggtgaaa gcagagattt aagccagtgc tagaactgat
6240

tggagtggga cagctgccca caatttagtt tgaaagacaa gttcagcaga tagattgaga
6300

gagaggaaag ctccctcaga ggaagtgatg tttgaacctg gccttgagaa attaatagaa
6360

atttgccaga tagaagtctg ccaccacacc tggctagttt ttgtattttt aatagagatg
6420

gggtttcact gttggccagg ctagcctcaa actcctggcc ttaagtcatc cacccgccta
6480

ggcctcacaa agtgttggga ttacacttgt gttgggtggc atgagccact gtgcctggcc
6540

aacttttaga tttttttttt ttgggagatg gagtctgtct cccaggctgg agtacagtgg
6600

tgctatcttg gctcactgca acctcagcct cctgagtagc tgggatcaaa ggcacccggc
6660

taattttttgt atttttagta gagacagggc tttcatcatg ttggccaggc tggtctcaaa
6720

ctcctgacct caggtgatcc acccacctcg gcctcccaaa gtgctgggat tacaggcgtg
6780
```

```
agccaccgcg cccagccaga aggtggcata tttatagcaa aggaaatagc atgtgttttg
6840

gttgatgtaa atatataagc tataacatgt agtgttctct ttagaacagt cgggtatgct
6900

gttacagttt tagataaatg tgaagcaaat gatgataaac tggatctgac tgactgtgct
6960

gagtctgttc aatccaaccc tgagcttcat gttctgtctc ttaacctcca aatagaccaa
7020

tgcccccatc aattccatcg ctcttctttc aggagcatgc aagttgtccc cttggacaaa
7080

cagatcacaa tcatagatag tccgagcttc atcgtatctc cacttaattc ctcctctgcg
7140

cttgctctgc gaagtccagc aagtattgaa gtagtaaaac cgatggaggc tgccagtgcc
7200

atcctttccc aggctgatgc tcgacaggta aaaggacccc ttctcatgag ctccttggag
7260

ccatcttctt tcatcataag cattttgagt agaaaaatct tggaagtgtt ttaaagtact
7320

ggcatgtcag atgaaggaca gctcctttgt ttggtttttt ttttaaggta gtactgaaat
7380

atactgtccc aggctacagg aattctctgg aattttttac tgtgcttgct cagagaagag
7440

gtatgcacca aaaggtgga atcccaaatg ttgaaggtgc tgccaaactg ctgtggtctg
7500

agtggacagg gtaagctttc ttttctgttg gcattttggt gaccactaga ataaaccttc
7560

ttttgacaca tcttattttt aatatcagtg cctcattagc ttactattgc catcccccta
7620

catcttggac tcctcctcca tattttaatg agagtattgt ggtagacatg aaaagcggct
7680

tcaatctgga agaactggaa aagaacaatg cacagagcat aagaggtgag aattgtgtgt
7740

cgctgctgtc ttcatcagct gacaggccag tggagctctt acctgtttac atgggcttgc
7800

tttctttccc agccatcaag ggccctcatt tggccaataG/C catccttttc cagtcttccg
7860

gtctgacaaa tggaataata gaagaaaagg acatacatga agaattgcca aaacggaaag
7920
```

```
aaaggaagca ggaggagagg gaggatgaca aagacagtga ccaggaaact gttgatgaag
7980

aagttgatgt aagtgtgtcc tccatgagtt aaaactgaag tgagttttct agcattataa
8040

tacataatgg aaggaactga agataggaaa tatttgaggc ttgtgatcca ttagccttaa
8100

ttttgcacat cccgttatat gtacctccaa agagttaatt tttcaggtac ataactactt
8160

ggattaaatg agcagacaag ggctactaat ccagcactat ttttctttgt cacacaggaa
8220

aacagctcag gcatgtttgc tgcagaagag acaggggagg cactgtctga ggagactaca
8280

gcaggtgagg caggcaaaag gggttctaac gaagcagcat ggtatagaat cacttttact
8340

ttttgaaaat ctctttattt tcctgcaata taggtgaaca gtctacaagg tcttttatct
8400

tggataaaat cattgaagag gatgatgctt atgacttcag tacagattat gtgtaacaga
8460

acaatggctt tttatgattt tttttttaac attttaagca gactgctaaa ctgttctctg
8520

tataagttat ggtatgcatg agctgtgtaa attttgtgaa tatgtattat attaaaacca
8580

ggcaacttgg aatccctaaa ttctgtaaaa agacaattca tctcattgtg agtggaagta
8640

gttatctgga ataaaaaaag aagataccta ttgaaaaatg taagttttat ttacagatca
8700

ggccacaggt tacaaaatta aaaccaacag cagttttgaa ttatctgtac cagctagctg
8760

aactagccat atcagttc
8778
```

FIGURE 2

Polymorphic Region	Sequence	SEQ ID NO
245 sense 5'	agcggaggcaggttg	2
245 antisense 5'	ggaagcacaaacaca	3
245 A variant	aggttgAtgtgtt	4
245 C variant	aggttgCtgtgtt	5
245 antisense T variant	aacacaTcaacct	6
245 antisense G variant	aacacaGcaacct	7
245 sense 3'	tgtgtttgtgcttcc	8
245 antisense 3'	caacctgcctccgct	9
1470 sense 5'	aggaggctaaaaagc	10
1470 antisense 5'	ggcttcttgtgaccc	11
1470 A variant	aaaagcAgggtca	12
1470 G variant	aaaagcGgggtca	13
1470 antisense T variant	tgacccTgctttt	14
1470 antisense C variant	tgacccCgctttt	15
1470 sense 3'	gggtcacaagaagcc	16
1470 antisense 3'	gcttttttagcctcct	17

FIGURE 3

Polymorphic Region	Primer Sequence	SEQ ID NO
245	ctcgtcagtggcttcagttc	18
245	ccttttcatattggctgtagaa	19
1470	aataggttcgagaacatcatcg	20
1470	ttggaactcctgggtctttc	21